REMARKS:

I. The Invention.

A. Why a table (or database) join operation will not accomplish the results of the invention.

The present invention is about two important features. First, a commonality is created, where none may have existed before, across two or more disparate databases, by creating qualitative and quantitative variables and recoding relevant data in the various databases. Second, using this new linking or commonality, the databases are combined, <u>not joined</u>, and a statistical cluster analysis is performed using data from each of the databases. Cluster analyses are also utilized to determine which of the variables provide greater discriminatory power as "statistical drivers". Only following these statistically significant cluster analyses of the combined data may conclusions be drawn, such as deriving a valid behavioral model. For example, data from one database may be used to predict behaviors of different individuals represented in a second database, or to overlay and describe other patterns and attributes of these individuals represented in the second database.

As discussed in the September 8, 2005 interview with the Examiner, the databases involved cannot be subject to a mere database table join operation, which is all the prior art has to offer. A mere table join of disparate databases, even if possible, would potentially provide horrible, statistically irrelevant and misleading results. Literally, garbage. Indeed, a join operation is *never* mentioned in the specification, and any prior art providing for table join operations for combining databases teaches away from the present invention.

For example, as disclosed in the specification, one of the databases may be a credit card transaction database, such as a Mastercard database (specification, p. 5, l. 3; p. 10. ll. 10 – 11), which includes millions of people, but has a limited depth of data, being confined to transactions at particular merchants or vendors, and being subject to confidentiality restrictions under federal regulations. Another database may be a Simmons National Survey database (specification, p. 10, ll. 13 – 14), with a much smaller number of individuals, such as in the thousands, but having a tremendous wealth and richness of data, with each database respondent essentially having provided answers to

thousands of questions, and with repeated surveys over time. In addition, each database typically contains information about different individuals (specification, p. 13, ll. 14 – 17).

The fact that a table join operation is not only inapplicable to the present invention, but also teaches away from the present invention, is readily apparent from the following example. If a table join were attempted, with person A from the Mastercard database being joined (based on some common attribute or variable) with data from person B from the Simmons National Survey database, without the present invention, there could be totally erroneous and <u>misleading results</u>. Continuing with the example, assuming a join operation based on a common variable of being high spenders (*e.g.*, as a quantitative variable corresponding to a qualitative variable of frugality), person A, who in accordance with the invention should belong in a first statistical cluster of high spending individuals with advanced college degrees who drive luxury sedans, attend the symphony and watch soap opera genre and cable news television programs, would be erroneously matched or joined into person B's second statistical cluster (of the present invention) of high spending individuals with some college education who drive sport utility vehicles, go to rock concerts and never listen to classical music, who watch police genre television programs, and who never watch the news.

Clearly, such a join operation would not provide valuable results. Indeed, the results would be misleading and useless – one could not validly predict person A's behaviors, likes, dislikes, etc., based on the data from person B, despite the join operation. Moreover, merely extending such a join operation to additional variables, without use of the present invention, would not be helpful, as it would only decrease the probability of finding suitable matches in the databases to join on the additional variables, to the point that no matches may result and leaving big holes and gaps in the supposed database integration.

In addition, such join operations would still not account for other, <u>non-matching variables</u>, which may be highly statistically significant. For example, supposing persons A and B can be joined on 3 matching variables (*e.g.*, spending, store selections, and vehicle selections); nonetheless, one may be a smoker and the other a non-smoker, one may be female and the other male, and one may be a professional engineer

and the other a professional musician, which are highly relevant disparities for innumerable purposes, such as for health and insurance marketing, for clothing marketing, and for magazine or professional journal marketing, for example.

Moreover, such join operations would not account for variances within the quantitative variables and distances between such variables, such as mean values of transactions of 10.5 versus 13.4, which would not be joined but nonetheless may be grouped in a cluster analysis and provide highly relevant and statistically accurate information.

As a consequence, without use of the present invention, there could be no guarantee that anything meaningful would result from a database integration based on a join operation; rather, the present invention is needed to fully and validly integrate the databases and obtain statistically significant results, including overlays of such information derived from these disparate databases. Essentially, the only way to account for all of this is to utilize the novel features of the present invention which, through the cluster analyses, can account for matching variables, non-matching variables, and variances and distances between variables, among other features.

Of course, depending upon the selected databases, it is possible that in some instances there is sufficient personally identifying information that there could be a match between some individuals who happen to be in both databases. For many databases, that is highly unlikely, particularly in credit databases where identifying information is held confidentially and is legally protected from disclosure. The claimed invention, however, is for methods and systems of integrating disparate databases, in a statistically significant manner, and creating behavioral models from the resulting integrated database, using data from each of the databases forming the integrated database. Such disparate databases, as indicated above, are fundamentally different from each other, and generally not amenable to a join or matching operation (*see, e.g.*, a typical dictionary definition of "disparate", which means "containing or made up of fundamentally different and often incongruous elements" (Webster's Ninth New Collegiate Dictionary, 1984)).

But even assuming for the sake of argument that, contrary to the claims, the databases are not disparate and that some matches could be made between individuals

known to be in both databases, such that their corresponding rows could be validly joined, *e.g.*, joining person A in database 1 to the same person A in database 2. Without use of the present invention, however, there would be absolutely no way to account for and combine or overlay data for any unmatched individuals, which could be a huge proportion of one or both databases. For example, supposing that 100% of a Simmons database of 10,000 individuals are matched into the Mastercard database having millions of individuals. Without the present invention, this leaves non-integrated and uncombined roughly 99.9% of the Mastercard database.

Without the present invention, these databases effectively remain non-integrated, as the prior art join operation does not provide any mechanism to combine the non-matching database members in a statistically significant and valid manner, let alone to create behavioral models from the resulting integrated database. Indeed, the only way for a join operation to succeed is trivial, where both databases contain 100% matching individuals (members), indicating that they are *100% non-disparate* and providing no real or effective result, as no new information and predictions result. Such a trivial operation misses the point of the invention, which is to provide a statistically significant way of porting or combining data concerning one set of individuals with data concerning a different set of individuals.

The present invention accomplishes this valid integration of these disparate databases first, by creating commonality, and second, by performing cluster analyses across the combined data, using data from each of the databases. Such commonality is created through identifying qualitative variables (Figure 3, step 200), expanding these variables into quantitative variables (Figure 3, step 210), and then converting the information in each of the databases based on the quantitative variables (Figure 3, step 225). In addition, various statistical analyses may be performed on the quantitative variables, such as a principal components analysis in which the variables are standardized, made to be "substantially orthogonal", and weighted (specification, p. 13, ll. 18 – 22). In addition, more significant quantitative variables may be selected as statistical drivers to provide higher discriminatory power, using the cluster analyses described below (specification, p. 15, ll. 7 – 13).

Once this commonality is then present in both databases, the databases may then be "combined" (Figure 3, step 230), a cluster analysis is performed (Figure 3, step 250), followed by generating a predictive, behavioral model (Figure 3, step 275). Independent claims 1 and 9 focus on this integration and formation of a behavioral model, while independent claims 17 and 30 focus on using the more statistically significant variables, referred to as statistical drivers (derived or selected from the other variables, claims 20, 21, 33, 34), the cluster analysis, and validating such a cluster solution.

B. The cluster analysis of the present invention.

The cluster analysis is explained in detail in the specification, including the use of iterative or repeated analyses to form clusters of members, and the selection of statistical drivers. In addition, the specification references software which, once the commonality has been created between disparate databases in accordance with the invention, can then be utilized to perform one or more cluster analyses.

First, as discussed in the specification with reference to Figures 3 and 6, candidate statistical drivers are initially selected from the quantitative variables ("blooming variables") as members with the most discriminatory power (specification, p. 15, ll. 9 – 12). These members are then "clustered into a preliminary cluster data set (step 510), and the discriminatory power of the clustered members is evaluated as a set according to, e.g., the root mean squared standard ("RMSSTD") statistic, and as an estimated R^2 for the model (step 520)" (specification, p. 15, ll. 17 – 20). Exemplary software which performs this statistical analysis is mentioned, such as FASTCLUS, which is known by those of skill in the art to perform a disjoint cluster analysis on the basis of Euclidean distances, as a k-means model (as described in its user manual, or obtained through a Google search of "FASTCLUS"). This process is repeated, with selection of "different members … as the candidates for the statistical drivers", until the discriminatory power of the cluster solution is satisfactory" (specification, p. 16, ll. 2 – 6). An "optimal number of clusters" is determined using "the estimated R^2 for the model, a cubic clustering criteria, the pseudo t statistics and the pseudo F statistics" (specification, p. 16, ll. 10 – 12). (An exemplary, first pass iteration of this process is further described on pages 16 -17 of the specification).

Next, an additional level of cluster analysis may also be performed, referred to as "superclusters", utilizing an additional statistical analysis referred to as "'Ward's' method", or the other R^2 and pseudo t and pseudo F statistics (specification, p. 17, ll. 8 – 16). In addition, unassigned members may be re-assigned using "a nearest neighbor strategy" by assigning "to the cluster whose centroid is nearest to the particular unassigned respondent in a multi-dimensional space" (specification, p. 18, ll. 3 – 5). Then, using additional "statistical summarization techniques", other behavioral characteristics which are not statistical drivers ("descriptors") can be used to further describe the corresponding clusters and provide additional information regarding the respondents in the integrated database (specification, p. 18, ll. 9 – 14).

This results in a "much more accurate and predictive model of the future consumer behavior" (specification, p. 19, ll. 6 – 7), and "a better estimation of the behavior of potential customers which are provided in the same clusters. For example, if the particular customers are assigned to the same cluster because they shopped in the department store identified in that cluster, the like to watch the same television show, they like to go to the movie theater on weekends, etc., it is significantly easier to predict that the behavior of these customers would be similar in other situations (e.g., where they travel on vacations, etc.)." (specification, p. 21, ll. 4 – 10).

As a consequence, it is readily apparent that the exemplary embodiments of the present invention utilize statistical analyses, such as cluster analyses, in order to combine disparate databases and provide corresponding behavioral modeling. The prior art does not disclose this novel methodology.

II. The Office Action.

A. In the Office Action mailed April 19, 2005, the Examiner made various objections to the specification and claim 19 (Office Action Points 4 and 5). The specification has been amended to address these typographical errors, and claim 19 has been cancelled.

B. In points 6 and 7 of the Office Action, a concern was raised as to whether qualitative variables and/or quantitative variables are common to the databases. As a consequence, the relevant portions of claims 1 and 9 have been amended to return to their original state, of quantitative variables being common to each of the databases, with corresponding amendments of claims 20 and 33. Claims 19 and 32 have been cancelled. In addition, it would be implicit that once a qualitative variable is determined to be in common, at least some of the resulting, derived quantitative variables from the qualitative variable would also be in common (*e.g.*, the "entries ... in the databases... are coded in terms of the blooming [quantitative] variables" (specification, p. 14, ll. 6 – 7)).

One potential source of confusion for what may or may not be a "qualitative variable which is common to" the databases, which may have resulted from earlier amendments or responses and should be clarified, is the difference between a database attribute (forming a column in a table) and a qualitative variable as used in the present invention and having a meaning known by those in the statistical fields. In many instances, a qualitative variable of the present invention may be an abstraction from a database attribute, which may or may not be common across the databases. For example, a "police genre television program" qualitative variable is an abstraction from different media database attributes, such as "watches CSI: Crime Scene Investigation" and "watches re-runs of NYPD Blue".

Similarly, using the examples of the specification, a qualitative variable "shops at Macy's" may be further abstracted into "shops at high-end department stores", in order to create a commonality with a second database which has "shops at Marshall Fields" (not a common qualitative variable), and further, to be a discriminating variable from "shops at Target" (as too many people may shop there, so it does not provide statistically discriminating information). As a consequence, there are instances in which a qualitative variable is derived or created in order to be common (create commonality) between two databases, as there may not be any common attribute data. Without this abstraction into a qualitative variable, there may not be any commonality between the attributes, and is an additional reason a simple join operation is inapplicable. This also is why, in the previous responses, some qualitative variables were viewed as not being in

common between the databases, until abstracted at this higher level or converted into quantitative variables.

C. Claims 34 and 36 have now been amended to correct the claim dependencies. As a consequence, the concern raised in points 8 and 9 of the Office Action have been corrected.

D. With regard to point 10 of the Office Action, as discussed in the interview, relevant elements of independent claims 1 and 17 have been amended to provide that they are accomplished by a processing device, such as a computer or microprocessor (specification, p. 10, l. 18 – p. 11, l. 14; Figure 2).

E. Various claims were rejected, under 35 USC Section 103(a), in various combinations, as being unpatentable over: (1) Gupta (V. R. Gupta, An Introduction to Data Warehousing, System Services Corporation, August, 1997 (Google)) ("Gupta") in view of Apte (C. Apte, Data Mining – an Industrial Research Perspective, IEEE Computational Science and Engineering, April-June 1997 (Google)) ("Apte") (Office Action, points 11 and 12); (2) Gupta and Apte in view of Anderson et al. U.S. Patent No. 5,974,396 ("Anderson") (Office Action, point 13); and (3) Gupta, Apte, and Anderson in view of claim 17 (Office Action, point 14).

As discussed during the interview, both Gupta and Apte provide only for table join operations, and teach away from the present invention. (Apte, paragraph 14 "extracted … and joined"; Gupta, section 2.2.4, de-normalization of data and table joins; section 2.4.1, multiple table joins to generate summary views.) Neither Gupta nor Apte disclose or suggest innumerable features of the various independent claims of the present invention, including without limitation, identification of qualitative variables, transformation into quantitative variables, combining (not joining) disparate databases based upon these new commonalities, performing a cluster analysis using data from the disparate databases, and behavioral modeling from the statistical (cluster) analysis. As a consequence, neither Gupta nor Apte, alone or in combination, disclose, suggest, or render obvious the claimed invention.

Also as discussed, Anderson discloses creating separate databases from a common pool of information, which are then linked trivially through pre-assigned unique identification numbers. The Anderson reference also teaches away from the present invention, and does not create "clusters" using statistical cluster analyses. Indeed, the use of "cluster" in Anderson is misleading; the supposed "clusters" in Anderson are just generic product groupings, determined in advance in order to save database space.

Anderson discloses a typical database system of a modern grocery store, in which customers are issued some type of "membership" cards, which uniquely identify a particular customer (Anderson, col. 4, ll. 19-33). As part of the card issuing process, the grocery store collects demographic information on the customer, which it tracks using an assigned member identification number ("MIN") (Anderson, col. 8, ll. 21-35). When a given customer shops, the card is scanned, so that information concerning what that individual purchased may be tracked using the assigned identification number (Anderson, col. 4, ll. 19-33; col. 7, ll. 22-41).

Rather than integrating disparate databases, the Anderson reference divides the information gathered into two different databases, which are linked ahead of time using the member identification number (Anderson, col. 4, ll. 19-33). With the assigned identification numbers, every piece of information is automatically matched to other data fields, such that there are no disparate databases in Anderson.

Because storing information for each product transaction may create an excessive volume of information (col. 3, ll. 20-29), using "predefined" product criteria, the Anderson reference assigns various products into *a priori*, generic product clusters, and also provides "predefined" consumer criteria for *a priori* consumer clusters (Anderson, col. 4, ll. 7-18). As an example of Anderson, a purchase of "cat food" would be pre-assigned in advance to a generic cluster of "pet food" (Anderson, col. 10, ll. 24-27). This information is then divided, not integrated, into separate database tables, for products and consumers, and linked using the pre-assigned identification number (*See, e.g.*, Figure 6 and col. 10, ll. 31-45). As all clusters are pre-defined, all clustering in Anderson is predetermined, in advance of any data analysis, and is not empirically derived from data analysis.

As a consequence, the Anderson reference does not provide any of the empirical clustering analyses and database integrations of the present invention. Rather, using predefined criteria for each type of consumer or product cluster, and by dividing information into separate databases linked in advance using identification numbers, Anderson also teaches away from the present invention.

Such teaching away is the antithesis of art suggesting that a person of ordinary skill go in the claimed direction. See *In re Fine*, 873 F.2d 1071 (Fed. Cir. 1988). This teaching away from Applicants' invention is a *per se* demonstration of lack of obviousness and a lack of anticipation.

In summary, the references do not disclose creating a combined, integrated database, from a plurality of *disparate* databases. The references do not disclose or suggest identifying qualitative variables in each database and converting them into quantitative variables to create grounds for commonality across the incongruous databases, and further, do not disclose converting the data within each database based on such quantitative variables.

In addition, other claimed features of the present invention are not disclosed or suggested in any of these references. For example, the references do not discloses creation of statistical drivers across each of the disparate databases, which are then used to recode and combine, not join, the various databases. Also for example, the references do not disclose or suggest creating a simultaneous cluster solution across such disparate databases, using information from each.

Moreover, the examiner has not presented any motivation, suggestion or teaching to combine these references. The mere fact that the references could be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680 (Fed. Cir. 1990). In addition, identification of any individual part claimed is insufficient to defeat patentability of the whole claimed invention. See *In re Kotzab*, 217 F.3d 1365 (Fed. Cir. 2000). Accordingly, no *prima facie* showing of potential obviousness has been made, and any assertions to the contrary have been clearly rebutted. *In re Rouffet*, 149 F.3d 1350 (Fed. Cir. 1998); *In re Mills*, supra. The rejection of claims as obvious under Section 103(a), therefore, should be withdrawn.

F. The additional cited references also do not disclose and do not suggest the claimed features of the present invention (Office Action, point 15):

(1) The Chaudhuri reference pertains to general data warehousing and online analytical processing (OLAP) and does not disclose the claimed features of the invention such as the creation of qualitative and quantitative variables, the combining of data from two or more disparate databases, the statistical, cluster analysis and behavioral modeling of the present invention;

(2) Fayyad et al. U.S. Patent No. 6,263,337 pertains to clustering across a singular database, and does not disclose the claimed features of the invention mentioned above;

(3) Fayyad et al. U.S. Patent No. 6,263,334 pertains to a nearest neighbor analysis across a singular database, and also does not disclose the claimed features of the invention mentioned above;

(4) The Farley reference pertains to meta-analysis, and also does not disclose the claimed features of the invention mentioned above; and

(5) Malloy et al. U.S. Patent No. 5,905,985 pertains to multi-dimensional data structures such as OLAP cubes, and also does not disclose the claimed features of the invention mentioned above.

As a consequence, the cited references, alone or in combination, do not disclose and do not suggest the claimed features of independent claims 1, 9, 17 and 30. The present invention, therefore, is not anticipated and is not rendered obvious by these references under Section 103, and the rejection of the claims should be withdrawn. In addition, because the remaining dependent claims incorporate by reference all of the limitations of the corresponding independent claims, all of the dependent claims are also allowable over the cited references.

On the basis of the above amendments and remarks, reconsideration and allowance of the application is believed to be warranted, and an early action toward that end is respectfully solicited. In addition, for any issues or concerns, the Examiner is invited to call or email the attorney for the applicants at the telephone number and email address provided below.

Respectfully submitted,

Max F. Kilger et al.

September 10, 2005

By

Nancy R. Gamburd
Attorney for Applicant
Registration No. 38,147
Phone: 312-876-0460
Cell: 312-399-9332
Cell: 408-429-3310
Fax: 312-276-4176
ngamburd@gamburdlaw.com

CERTIFICATE OF FACSIMILE TRANSMISSION

I hereby certify that the within and foregoing AMENDMENT AND RESPONSE UNDER 37 CFR 1.111 AND 1.115 (27 pages), TRANSMITTAL (PTO/SB/21), FEE TRANSMITTAL (PTO/SB/17), PETITION FOR EXTENSION OF TIME (PTO/SB/22, 2 copies), and CHANGE OF CORRESPONDENCE ADDRESS (PTO/SB/122) for Max F. Kilger et al., Serial No. 09/610,704, entitled "Process and System for Integrating Information from Disparate Databases for Purposes of Predicting Consumer Behavior", have been transmitted via Facsimile to 571-273-8300 addressed to Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on September 10, 2005.

Nancy R. Gamburd
Reg. No. 38,147